

Against the Grain

Volume 22 | Issue 5

Article 27

November 2010

Biz of Acq-Where's Web?

Aline Soules

California State University, soulesae@gmail.com

Michelle Flinchbaugh

University of Maryland Baltimore County, flinchba@umbc.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Soules, Aline and Flinchbaugh, Michelle (2010) "Biz of Acq-Where's Web?," *Against the Grain*: Vol. 22: Iss. 5, Article 27.

DOI: <https://doi.org/10.7771/2380-176X.5660>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Biz of Acq — Where's Web?

by **Aline Soules** (Librarian and Professor, University Libraries, California State University, East Bay, 25800 Carlos Bee Blvd., Hayward, CA 94542; Phone: 510-885-4596; Fax: 510-885-4209) <soulesae@gmail.com>

Column Editor: **Michelle Flinchbaugh** (Acquisitions Librarian, Albin O. Kuhn Library, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250; Phone: 410-455-6754; Fax: 410-455-1598) <flinchba@umbc.edu>

Column Editor's Note: *Aline Soules provides this interesting article on considerations for handling free Web resources in the library. I was delighted to receive this article because the decision to add, or not add, free Web resources to the library collection is one that I've been dealing with in the course of my ordinary acquisitions work for some time. At my institution, Acquisitions began receiving occasional requests for items available for free on the Web, and lacking a policy or philosophy, we began determining on a case-by-case basis what to do with those requests. If we think the item requested would likely remain on the Web on a long-term basis, we catalog the site. In other instances, where we suspect the material will be removed, such as television news streaming video of political debates, we recommend that the faculty member add the resources to courseware instead. Eventually we will need to develop a carefully considered collection development policy for free materials, and this article brings up many of the issues that we'll need to discuss. — MF*

Introduction

We're all familiar with games like *Where's Waldo* or *I Spy*. The only reason it's tricky to find what you're looking for is that it's obscured by everything around it. Similarly, it's easy to enter a term on the Open Web and get a million results, but will you find what you're looking for?

If you want facts, the answer is probably yes, but what about searching for Websites relevant to your scholarly interests without also retrieving far too many extraneous results? The ability to enter your own natural terms and get back exactly what you want is far from perfected yet. Integration with other types of materials is possible if the Open Web is included in federated searching tools and you search through those tools, but not guaranteed if a search is initiated on the Open Web. If you *do* find something useful, what is the best way to save it? Simply, can the Open Web really give you what you want?

Where are Websites these days? Not just on the Internet. Websites are also:

- indexed in commercial databases
- cataloged in online catalogs
- gathered on subject guide pages by librarians and in various forms by users themselves
- developed by groups with like interests
- referenced on **Facebook**, **Twitter**, and **YouTube**
- affiliated with specific books, such as supplements to textbooks

- replicated in print and electronic books
- linked from anywhere to anywhere

The good part about this is that these practices follow **Weinberger's** principle to "put each leaf on as many branches as possible,"¹ the theory being that these sites will be more easily and more often found.

Searching for a Web Site

But is this always the case? *The Walt Whitman Archive*,² for example, is a noted Website of primary sources (it even provides access to a recording of what is thought to be Whitman's voice). It is accessible through the Open Web, but also indexed in the *MLA International Bibliography*, linked from numerous other Websites, and included in some libraries' online catalogs. If you know to look for "The Walt Whitman Archive," you'll find it easily through any of these sources; however, unless you know that specific title, you are more likely to search "Walt Whitman" or even "Whitman," as so many of my students do. If you enter Walt Whitman as a search phrase in these various sources, what do you find? I conducted searches in these sources on July 18, 2010.

The Archive showed up as the second entry in a **Google** search resulting in about 2,650,000 results. It appeared right after the Wikipedia entry (which will users pick first?). In **Google Scholar**, it did not appear in the first three screens (30 entries) of the 43,600 results. Will users go past three screens? As far as three screens?

In the *MLA International Bibliography* (searched on the **EBSCOhost** platform), it did not appear in the first three screens (30 entries) of the 3,835 "all results." On looking at the "source types" list on the left

navigation bar, I had to click "more" before the category "Websites" appeared as an option in this particular search. On clicking "Websites," the *Whitman Archive* was the first entry of 21 results culled from the original 3,835. Having this particular archive (there are others) indexed in these various places, therefore, doesn't necessarily increase the likelihood that users will find it.

As for the catalog, **California State University, East Bay** does not have an entry for the Archive, but a search in the **LINK+** consortium of 51 academic and public libraries to which **Cal State East Bay** belongs yielded 2,795 results and it was not in the first screen of 50 results, although a search for "Walt Whit-

man Archive" yielded 5 results, of which this specific archive was the second one listed. In other words, this Archive may be included in all these sources, but it is quickly buried among the large results set of a simple author search, even if the search includes both the first and last names of the author. This leads to the question of whether it is helpful to include the Archive in all these sources or whether including it causes the user to be confronted with too many results overall, with Websites (and potentially other useful items) getting lost in the overall list. Further, in spite of the fact that the **Wikipedia** entry came up first on an Open Web search, the Archive was prominently displayed as the second choice and, in the end, proved to be the most effective in drawing this source to my attention.

To Collect or Not to Collect?

The first decision libraries face is whether they will collect Websites at all. If I were to base such a decision on my one experience with the *Whitman Archive*, the answer would be quite clear, but a real study of many Websites is required to understand the full impact of such searches. If librarians do decide to collect Websites, what policies should they establish for collecting them and including them in various locations? Sites may be collected for their subject suitability, but there are other considerations. For example, if the general criteria for selection include the anticipated number of users, how is that assessed for a Website before it is collected? Also, once collected, how can currency be assessed long term? Although a book has a built in age, its aging process is predictable. Not so with a Website. Furthermore, while a Website like the *Whitman Archive* doesn't have a cost attached to it (donations are accepted), there are acquisitions and budgetary implications in

terms of staff time, record maintenance, etc. If a decision is made to create a record for a Website, is it better to catalog it or provide a link through selectors' discipline-specific subject guides or provide it as part of a separate set of library Web pages?

Library Created Web Sites

In addition to collecting some Websites, librarians are putting their own collections of information on Websites, which are subsequently cataloged. One example of an individual contribution is **Bailey's Scholarly Electronic Publishing Bibliography**,³ a free eBook that includes over 3,600 references to

continued on page 67



works about scholarly electronic publishing. If you are on an appropriate listserv, you will get a monthly notification that the bibliography has been updated. The 2008 annual edition of this bibliography, however, was published as a print book, also available on **Kindle**, both for a cost. The 2009 annual edition also appears in print as part of *Digital Scholarship 2009*, another book in traditional print format, this time with no **Kindle** version, evoking an interesting sense of going backwards.

Libraries are also engaged in extensive digitization, much of which ends up on Websites. Titles like **Charles J. Kappler's** "Indian Affairs: Laws and Treaties" at Oklahoma State,⁴ first published in 1903-1904 by the U.S. Government Printing Office, are also linked from other Websites. **Kappler's** work is linked, for example, from one page in **Wunder** and **Steinke's History Matters** Website, designed for high school and college teachers and students. *History Matters* is "a gateway to Web resources and offers other useful materials for teaching U.S. history."⁵ If a researcher wants to read more about **Kappler's** documents, there is also a paper by **Bernholz** and **Holcombe** that is available on another Website,⁶ dated Feb. 19, 2005, as well as through *Science Direct*,⁷ where it is indexed because it was published in *Library Collections, Acquisitions, and Technical Services* in March, 2005. Note that the Website was available earlier than the published paper.

User Response

Users, therefore, have many opportunities to find these resources (the "leaves on many branches" concept), but if they stumble first on the *Science Direct* listing and their library doesn't have access to the publication, what happens next? Do they give up? Ask for it through Interlibrary Loan, in which case, someone must expend time to identify the available free version? Or, worst case scenario, get a pop-up Web screen that asks them to pay for access? This intricate availability of "leaves on many branches" may not always be the blessing **Weinberger** envisions, especially with users who may or may not fully understand where these leaves originate, who don't automatically question which are fee and which are free, and who don't understand which are available with a few simple clicks and which are restricted through passwords, payment requirements, etc.

The Next Web

While this multiplicity of access will likely continue for some time in this transitional period, some librarians and other researchers are taking a different approach to finding information on the Web — the Semantic Web. As far back as 1998, **Tim Berners-Lee** provided a roadmap for the Semantic Web and wrote, "This document is a plan for achieving a set of connected applications for data on the Web in such a way as to form a consistent logical Web of data (Semantic Web)."⁸ If machines can understand the meaning of information on the Web through metadata, it is theorized that they can deliver more related information to the user. Between then and now,

The indispensable reference for all who work with words

The Chicago Manual of Style

SIXTEENTH EDITION

"The new go-to publishing manual is here! Responding to the demands made of publishing professionals in the digital age, the 16th edition continues to provide information for word lovers but has enhanced its guidelines to include electronic workflow and processes. . . . This venerable bible also has a new pastel blue bookjacket to alert users to the style, usage, and technology updates within, though the familiar orange cloth lies below. **A worthy, welcome addition to every library collection** as well as professional wordsmiths and educated readers of all persuasions."

Library Journal

1008 pp., 4 halftones, 54 line drawings, 15 figures, 12 tables, 1 musical example

ISBN: 978-0-226-10420-1 • Cloth \$65.00

The Chicago Manual of Style Online is available to libraries
by subscription. For rates and information, go to
www.chicagomanualofstyle.org. Free institutional trials are available.

The University of Chicago Press

development of the Semantic Web has continued through Berners-Lee and the World Wide Web Consortium (W3C), of which **Berners-Lee** is Founder and Director. W3C, whose "mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web,"⁹ envisions "one" Web, a Web of linked data using technologies focused on linking strategies, query languages, and vocabularies.

This effort has resulted in a plethora of initiatives related to interoperability efforts, vocabulary explorations, and ontologies, but the key question is whether these ef-

forts will result in more effective results or just more results, at least for the foreseeable future. The purpose of searching any source or group of sources is to gather the relevant information for the question being searched while keeping away what is irrelevant. If we end up with "one" Web, i.e., one place to search, what will the results give us? Will we receive duplicate results, only to find that we've chosen the wrong source for the actual data/full-text because it costs or because it's not the latest version? Will we get one result with many options for seeking the data/full-text, but no indication of which one is most

continued on page 68

appropriate for our situation? Will we even find what we're looking for? Will we have so many results that we don't go past the first few screens and, despite "relevance" ranking, lose valuable possibilities or not find what we're looking for at all?

Discovery

The answer may now lie in "discovery" services. The problem with commercial options, however, is that in order to ensure that users stay inside their discovery product, vendors include many pieces of information that might be better if excluded, e.g., Websites. In the **California State University** system, however, **David Walker** has developed an open source interface to **Metalib** by **ExLibris**. The possibilities can be seen in presentations **Walker** has made available on the Web.¹⁰ While the traditional concept of a catalog as the centerpiece of library holdings has been shifting for some time, this really makes the catalog one piece in the puzzle and a small one at that. A library could include the Open Web as a search option, along with the catalog, databases, or any other source, all on an equal footing. The implication is that Websites do not have to be collected, cataloged, or acquired in separate transactions. They can simply be searched along with everything else.

The positive aspect of this idea is that librarians can sidestep difficult decisions about what Websites to collect or not collect for inclusion in the catalog or subject guides. The Open Web results would show up as part of the results set. The negative aspect of this is, once again, the size of the results set, unless the user inserted specific and, probably, multiple terms. In the discussion portion of a Webinar on Xerxes 2.0,¹¹ **Walker** indicated that users are expected to "interface" with their results. That suggests that they will refine their searches, think of other terms, add terms, etc. But do they? Users, particularly the students with whom I work, often use terms that are far too broad, e.g., "I have to write a paper on women," after which they enter "women" as their search term. Users are also used to looking at the first one or two screens of results and either picking something or deciding that there's nothing there, if the results don't mesh exactly with what they think they need.

Full Circle

This type of behavior is what leads librarians to "collect" Web information in the first place, whether in their catalogs, on their subject pages, or elsewhere, in order to try to help users find key resources. If the searches for the *Whitman Archive* are any indication, however, the most successful search for the Archive was actually on the Open Web, not in the library catalogs or even **Google Scholar**. Perhaps it's time to re-think the idea of collecting Websites altogether. This brings us back, full circle, to the question of whether some Websites should or should not be collected and drawn to users' attention as part of the collection of scholarly materials available.

Another alternative is to create a multi-tiered approach and make that available through the discovery tool. Currently, users can specify in database searches whether they want to search a specific type of material, such as scholarly or peer-reviewed items only or just articles. This is not presented preferentially, however. It's simply a choice of equal value. Students who are told by their professors that they must have three scholarly articles, for example, will introduce that limit to their searches, if they are aware of it, in order to meet the requirement. An alternate possibility is to structure a discovery tool to search categories chosen by librarians first, then, in a second tier, offer them the Open Web. Some might argue that this is a role that librarians shouldn't play, but we already play that role by what we collect in the first place.

Conclusion

The Open Web has useful information. To find it, we need the following:

- continued development of discovery tools, which are in their early stages
- willingness to recognize the value of the Open Web to the point where we include it in discovery tools
- increased attempts to help users understand the meaning of their results list, and
- concerted efforts to encourage users to interface with and manipulate results to better effect rather than just picking the first items on the list

That is where we should put our efforts rather than trying to compensate for this transitional period by collecting Websites, creating records, and expanding library Web pages, all of which make more work for ourselves than we can ultimately manage. 🍷

Endnotes

1. **Weinberger, David**. *Everything is miscellaneous: The power of the new digital disorder*. New York: Henry Holt, 2007.
2. **Folsom, Ed, and Kenneth M. Price**, eds. *The Walt Whitman Archive*. Lincoln, NE: Center for Digital Research in the Humanities at the University of Nebraska-Lincoln, 2010. <http://www.whitmanarchive.org/> (accessed July 18, 2010).
3. **Bailey, Charles W**. *Scholarly Electronic Publishing Bibliography*. Houston: Digital Scholarship 1996-2010.
4. **Kappler, Charles J.**, Comp. *Indian Affairs: Laws and Treaties*. Stillwater, OK: Oklahoma State University, n.d. <http://digital.library.okstate.edu/kappler/> (accessed July 25, 2010). This work covers U.S. Government treaties with Native Americans from 1778-1883 (Volume II) and U.S. laws and executive orders concerning Native Americans from 1871-1970 (Volumes I, III-VII).
5. **Wunder, John R., and Christopher Steinke**. *History Matters: The U.S. Survey Course on the Web*. Lincoln, NE: University of Nebraska, n.d. <http://historymatters.gmu.edu/d/1056/> (accessed July 25, 2010).
6. **Bernholz, Charles D., and Suzanne L. Holcombe**. *The Charles J. Kappler 'Indian Affairs: Laws and Treaties' Internet Site at the Oklahoma State University*. Lincoln, NE: University of Nebraska, Feb. 19, 2005. <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1009&context=libraryscience> (accessed July 25, 2010).
7. **Bernhold, Charles D., and Suzanne L. Holcombe**. *The Charles J. Kappler 'Indian Affairs: Laws and Treaties' Internet Site at the Oklahoma State University. Library Collections, Acquisitions, and Technical Services* 29:1 (March 2005), pp. 82-89; doi 10.1016/j.lcats.2005.01.001. <http://0-www.sciencedirect.com/science/journal/14649055> (NOT accessed July 25, 2010).
8. **Berners-Lee, Tim**. *Semantic Web Road Map*. Oct. 14, 1998. <http://www.w3.org/DesignIssues/Semantic.html> (accessed July 25, 2010).
9. World Wide Web Consortium. *W3C Mission*. 2009. <http://www.w3.org/Consortium/mission.html> (accessed July 25, 2010).
10. **Walker, David**. *Thoughts, Resume, Presentations*. 2010. <http://library.calstate.edu/walker/> (accessed July 31, 2010).
11. **Walker, David**. *Xerxes 2.0*. Webinar. July 28, 2010. This Webinar was given to **California State University** users and is not available on the Open Web at the time of writing.

Porter's famous play. It seemed appropriate in the times we find ourselves in right now. And we have tried to stir up the program this year in the spirit of **Anything Goes!** We have **Tech Talks** on Saturday morning which will offer publishers and vendors 30 minutes each to demonstrate their most innovative products during concurrent sessions. There will even

be food to keep us alert! The fabulous **Beth Bernhardt** and **Leah Hinds** (could call them the Minnesota Twins except neither one of them is from Minnesota, oh well) have cooked up the idea of **Happy Hour concurrent sessions** this year! Sounds like a unique idea, yes?

"Some good news for those in the book industry: the amount of time spent reading books remained unchanged in the quarter, with just under 60% saying they read books daily or weekly." Milliot's Book Consumer Annual Review, p.19.

Have to say that we were literally bombarded with **great proposals** for the **Conference** this year! That's just one of many reasons the program is full, full, full!

We are going to have a different kind of opening this year as well. The enterprising and

continued on page 77